Monte Carlo Exact Goodness-of-fit Tests for Nonhomogeneous Poisson Processes

Bo H Lindqvist

Department of Mathematical Sciences Norwegian University of Science and Technology N-7491 Trondheim, Norway

> Bjarte Rannestad Lyse Energi AS Breiflåtveien 18, Mariero N-4017 Stavanger, Norway

Abstract

Nonhomogeneous Poisson processes (NHPPs) are often used to model failure data from repairable systems, and there is thus a need to check model fit for such models. We study the problem of obtaining exact goodness-of-fit tests for parametric NHPPs. The idea is to use conditional tests given a sufficient statistic under the null hypothesis model. The tests are performed by simulating conditional samples given the sufficient statistic. Algorithms are presented for testing goodness-offit for the power law and the log-linear law NHPP models. It is noted that while exact algorithms for the power law case are well known in the literature, the availability of such algorithms for the log-linear case seems to be less known. A data example, as well as simulations, are considered.

Keywords: Sufficiency, Conditional test, Gibbs sampling, Power law NHPP, Log-linear law NHPP

1 Introduction

As a motivating example, consider the following data from a reliability growth program, taken from Leitch [1] (and analyzed in Section 7 in the present

paper). For a particular repairable system are observed n = 10 failures, at times 103, 315, 801, 1183, 1345, 2957, 3909, 5702, 7261, 8245. It is of interest to have information on the random process behind the data. The reasons for this could be, for example, to be able to predict future failures, or to obtain increased knowledge of the underlying phenomenon. For certain applications the motivation could otherwise be to compute optimal maintenance intervals, or in software reliability applications to predict when to stop debugging a program.

Nonhomogenous Poisson processes (NHPP) are widely used as models for such failures of a repairable system. In practice it may thus be of interest to check the NHPP property for a given set of data, by performing a so called statistical goodness-of-fit test.

A nice and informative discussion of how to perform goodness-of-fit testing in NHPP models is given in Baker [2]. In common use are general tests based on the Cramer-von-Mises test or the Kolmogorov-Smirnoff test (Bain and Engelhardt [3], Crow [4], Park and Kim [5]), but as Baker [2] advocates, there is often a need for "purpose-built" tests to detect special types of departures from NHPPs. This includes, for example, tests with power against processes for which the NHPP property itself is violated, and not necessarily just the functional form of the intensity.

One way to derive such tests is to embed the models in more general parametric failure models and use standard likelihood ratio tests (see e.g. Lindqvist et al. [6]). A practical problem with such an approach, and also with most standard goodness-of-fit tests, is that the observed number of failures is often too small to justify the use of asymptotic distributions for computation of critical values or p-values. This applies in particular to the example considered above, and we shall see later (Section 7) that asymptotic tests may be inappropriate here.

The above discussion motivates the need for exact goodness-of-fit tests when samples are small. The advantage of exact tests is that the nominal significance level holds exactly (possibly modulo a Monte Carlo error), while this is not necessarily the case for asymptotic tests. It is well known that the power law NHPP (see Section 2 for definition) admits exact tests. Indeed (Bain and Engelhardt [3], Baker [2]), the estimators of the parameters have a certain pivotal property which can be utilized in computing distributions of test statistics. This pivotal property was utilized in goodness-of-fit testing in for example [2] and [5]. Goodness-of-fit testing in power law NHPPs has, furthermore, been considered in a large number of articles, for example Kumar and Klefsjö [7], Gaudoin [8, 9], Crétois et al. [10], Gaudoin et al. [11], Zhao and Wang [12]

The aim of the present study is to consider in some generality, how one

can obtain exact goodness-of-fit tests for parametric NHPP models by conditioning on a sufficient statistic under the null model. While the case of power law NHPP is well studied in the literature, as already pointed out, we shall still consider it here as a reference case. A main purpose of our paper is, on the other hand, to study how similar analyses can be done for the log-linear law NHPP, which seems to be the standard parametric "competitor" to the power law model in reliability studies. It turns out, however, that exact analyses are not that straightforward for this model because a pivotal structure like the one for the power law is not present. This fact was apparently first noted by Lee [13], who considered conditional testing given a sufficient statistic in the log-linear model, but was only able to give an approximate solution. An interesting exception here is the exact test presented by Gaudoin [9]. This test, which is probably the first exact test presented for the log-linear NHPP, is based on the conditional probability integral transformation (CPIT) and utilizes conditioning on the appropriate sufficient statistic. Gaudoin [8] derives similarly CPIT tests for the power law model. As noted in [9], the tests based on CPIT are often not so powerful, however. Zhao and Wang [12] consider goodness-of-fit testing in particular for the log-linear model, but use asymptotic theory for maximum likelihood and hence do not obtain exact tests. The problem connected to such tests is briefly illustrated in a simulation in Section 7.

The tests considered in this paper are all based on Monte Carlo simulations of the relevant conditional distributions. They are thus exact only up to simulation error. The tests may therefore be called Monte Carlo exact. However, in principle such errors may be made as small as desired by increasing the number of simulations so the notion of exact testing is still warranted.

It should be stressed that our emphasis is on how to obtain exact tests when a goodness-of-fit statistic Z is given. The construction or choice of test statistics Z is hence beyond the primary scope of this article. Still we discuss the problem and illustrate by numerical studies.

The paper is organized as follows. Section 2 contains the basic setup and assumptions. In particular we present likelihood functions and sufficient statistics of the two parametric models under particular study. In Section 3 we describe the conditional tests in general terms, while in Section 4 we present algorithms for obtaining the needed conditional samples from the two parametric models. Some popular goodness-of-fit statistics for NHPPs are presented in Section 5. The modifications to be done when going from time censoring to failure censoring are summarized in Section 6. Section 7 contains a data example and some simulations, while the last section, Section 8, contains a discussion of results and possible extensions.

2 Models, null hypotheses and sufficient statistics

In analyses of repairable systems one traditionally considers either failure censoring, meaning that the process is observed from time t = 0 and until a prespecified number n of failures have occurred; or time censoring, where the process is observed until a given time τ . To save space, we will mainly assume time censoring in our theoretical derivations, and then discuss the necessary modifications needed to handle failure censoring in a separate section (Section 6). It is believed that time censoring is of most practical interest, although laboratory tests may be set up as experiments which run until a predescribed number of failures have occurred.

Thus, suppose we observe an NHPP with intensity function $\lambda(t)$ in a time interval $[0, \tau]$. Let $N(\tau)$ denote the number of failures occurring in the interval, and let the failure times be denoted $T_1, \ldots, T_{N(\tau)}$. The data can thus be written in the form

$$T = (N(\tau), T_1, \dots, T_{N(\tau)}), \tag{1}$$

and the log-likelihood function resulting from these observations is given by (Crowder et al. [14, p. 66], Meeker and Escobar [15, p. 413]),

$$\sum_{j=1}^{N(\tau)} \log \lambda(T_j) - \int_0^\tau \lambda(u) du.$$
(2)

In the present paper we shall consider null hypotheses of the form

$$H_0: T \sim \text{NHPP}(\lambda(\cdot; \theta)) \text{ vs. } H_1: H_0 \text{ does not hold},$$
 (3)

where by NHPP($\lambda(\cdot; \theta)$) we mean an NHPP with intensity function $\lambda(\cdot; \theta)$ for some θ in the appropriate parameter space. The null hypothesis is just the statement that the observations T come from this family of NHPP models. Any test of H_0 vs. H_1 can now be called a goodness-of-fit test.

The two most popular parametrizations of NHPPs are the *power law*, with intensity function given by

$$\lambda_p(t) = abt^{b-1}; \ a, b > 0, \ t > 0,$$

and the *log-linear law*, with intensity function

$$\lambda_{\ell}(t) = \exp(a + bt); \quad -\infty < a, b < \infty, \ t > 0.$$

These will serve as our main examples in the present paper.

For the power law NHPP the log-likelihood function (2) becomes

$$l_p(a,b) = N(\tau)(\log a + \log b) + (b-1)\sum_{j=1}^{N(\tau)} \log T_j - a\tau^b.$$
 (4)

The sufficient statistic is hence, by the factorization theorem (Casella and Berger [16]),

$$S_p(T) = (N(\tau), \sum_{j=1}^{N(\tau)} \log T_j).$$
 (5)

Sufficiency means, intuitively, that this statistic contains all the information about the parameters that is contained in the data. In particular, looking at (4), it is clear that the maximum likelihood estimators of the parameters depend on the data only through this statistic. The usual definition, and the key feature to be used in this paper, is that a statistic is called sufficient if the conditional distribution of the data, given the value of the sufficient statistic, does not depend on the parameters (Casella and Berger [16]). The factorization theorem, as cited above, is usually the simplest way of finding a sufficient statistic.

For the log-linear law NHPP we get, using (2), the log-likelihood function

$$l_{\ell}(a,b) = N(\tau)a + b \sum_{j=1}^{N(\tau)} T_j - (e^a/b)(e^{b\tau} - 1).$$
(6)

Hence the sufficient statistic is

$$S_{\ell}(T) = (N(\tau), \sum_{j=1}^{N(\tau)} T_j).$$
(7)

3 Conditional testing given the sufficient statistic

Suppose that we have given the observation T on the form (1). For a given parametric family $\{\lambda(\cdot, \theta)\}$ we are interested in testing the null hypothesis H_0 in (3).

Let there be given a test statistic $Z \equiv Z(T)$, which is a function of the data T, and which is a goodness-of-fit statistic in the sense that it is designed to have the ability to reveal departure from the null hypothesis.

Let $S \equiv S(T)$ be the sufficient statistic for the unknown parameter θ under the null hypothesis model, provided such one exists. This sufficient statistic is generally multivariate, usually with the same dimension as the parameter θ (see for example (5) and (7)). Suppose also, for simplicity, that large values of Z correspond to violation of the null hypothesis.

An α -level conditional test of H_0 rejects, conditionally given S = s, when $Z \ge k(s)$, where k(s) is a critical value chosen such that

$$P_{H_0}(Z \ge k(s)|S=s) = \alpha. \tag{8}$$

(We assume for simplicity that exact equality to the right of (8) is always possible, which will be the case in our applications). In order to find the critical value k(s) we need to know the conditional distribution of Z given S = s. Since this distribution, by sufficiency, is independent of the unknown parameters under the null hypothesis model, it can at least in principle be found and then applied to calculate k(s). Of course, this might in practice be difficult or even impossible to do analytically, and that is why we shall use Monte Carlo simulations in this paper.

For now, suppose we are able to simulate realizations of the conditional distribution of Z given S = s. Let the observed data be T_{obs} (in the form (1)), and let $s_{obs} = S(T_{obs})$ and $z_{obs} = Z(T_{obs})$ be the observed values of, respectively, the sufficient statistic and the test statistic Z. Instead of computing the value of $k(s_{obs})$ we shall consider the conditional *p*-value,

$$p_{obs} = P_{H_0}(Z \ge z_{obs}|S = s_{obs}). \tag{9}$$

Observe that (8) and (9) together imply that $p_{obs} \leq \alpha$ if and only if $z_{obs} \geq k(s_{obs})$. Hence the α -level conditional test is equivalent to the test which rejects the null hypothesis if $p_{obs} \leq \alpha$. The key to perform the conditional test is hence to compute the conditional p-value in (9).

For this, we simulate a large number, M say, of realizations T^* from the conditional distribution of T given $S = s_{obs}$ and compute $Z^* = Z(T^*)$ for each of them. The conditional *p*-value p_{obs} is then approximated by the relative frequency of the event $Z^* \geq z_{obs}$, i.e.

$$\hat{p}_{obs} = \#\{Z^* \ge z_{obs}\}/M.$$

It is clear that \hat{p}_{obs} can be made arbitrarily close to p_{obs} by using a sufficiently large M.

Note that a conditional test as described above is also an exact unconditional α -level test, since we have

$$P_{H_0}(\text{reject } H_0) = E_S[P_{H_0}(\text{reject } H_0|S)] = E_S[P_{H_0}(Z \ge k(S)|S)] = \alpha,$$

by definition of the function k(s). However, in practice we have a Monte Carlo exact test due to the approximation \hat{p}_{obs} .

4 Conditional sampling for parametric NHPP models

In the present section we give recipes for simulating samples of the form (1) conditionally given the appropriate sufficient statistic, needed for the computation of conditional *p*-values. We do this in detail for the two featured NHPP models, power law and log-linear law. A brief discussion of how to do this for more general models is given in the final section of the paper.

A classical trick for computation of conditional distributions given the sufficient statistic is to choose parameter values which give rise to particularly simple models. This can be done since the conditional distribution is the same for all parameter values. Indeed, for the power law and the log-linear law we shall use parameter values corresponding to homogeneous Poisson processes with unit intensity (denoted HPP(1)).

This means that the desired samples T^* can be found as samples from a HPP(1), conditionally given the sufficient statistic. Note that the statistics (5) and (7) include the number of failures, $N(\tau)$ as one component. Our further clue is now to condition on $N(\tau) = n$ first.

It is well known (e.g. Ross [17]) that conditionally on $N(\tau) = n$, the event times of an HPP(1) are distributed as the order statistics of n i.i.d. random variables uniformly distributed on $[0, \tau]$. In the following we denote this distribution by $U[0, \tau]$.

Thus, suppose our data are $T = (n, T_1, \ldots, T_n)$, that is we have observed $N(\tau) = n$ events in $[0, \tau]$. Let, further, $\sum_{j=1}^n \log T_j = s_1$ and $\sum_{j=1}^n T_j = s_2$. Then, the simulation of new samples $T^* = (N^*(\tau), T_1^*, \ldots, T_{N^*(\tau)})$ conditional on the sufficient statistics in the power law and log-linear cases, reduces to

- (i) simulate *n* i.i.d. values X_1, \ldots, X_n from the distribution $U[0, \tau]$, conditional on respectively $\sum_{j=1}^n \log X_j = s_1$ and $\sum_{j=1}^n X_j = s_2$,
- (ii) put $N^*(\tau) = n$, $T_j^* = X_{(j)}$ for j = 1, 2, ..., n, where $X_{(1)}, ..., X_{(n)}$ is the increasing ordering of the sample $X_1, ..., X_n$.

The simulation problem is hence transformed to a problem of conditional sampling from uniform variables. It is the purpose of the next two subsections to show how this can be done in the two featured cases.

4.1 The power law case

As indicated in the Introduction, the power law case admits well-known simple exact algorithms for simulating from the conditional distributions. The following lemma shows how to perform the task (i) given above. We write s instead of s_1 for simplicity.

Lemma 1 Let X_1, \ldots, X_n be independent and uniformly distributed on $[0, \tau]$. Then samples from the conditional distribution of X_1, \ldots, X_n given $\sum_{i=1}^n \log X_i = s$ can be simulated by the recipe

- 1. Draw U_1, \ldots, U_n i.i.d. from U[0, 1].
- 2. Let $X_i = \tau U_i^{(s-n\log \tau)/\sum_{i=1}^n \log U_i}$.

Proof: Let U_1, \ldots, U_n be i.i.d. from U[0, 1]. It is well known that $-\log U_i$ is exponentially distributed with expected value 1. Hence the vector

$$\left(\log U_1 / \sum_{j=1}^n \log U_j, \dots, \log U_n / \sum_{j=1}^n \log U_j\right)$$

is independent of $\sum_{j=1}^{n} \log U_j$. A standard proof of this uses Basu's theorem (Casella and Berger [16]). Using this independence result we can write

$$P(u_1 \le U_1 \le u_1 + \Delta_1, \dots, u_n \le U_n \le u_n + \Delta_n | \sum_{j=1}^n \log U_j = t)$$

$$= P(\frac{\log u_1}{t} \le \frac{\log U_1}{\sum_{j=1}^n \log U_j} \le \frac{\log(u_1 + \Delta_1)}{t}, \dots, \frac{\log u_n}{t} \le \frac{\log U_n}{\sum_{j=1}^n \log U_j} \le \frac{\log(u_n + \Delta_n)}{t})$$
$$= P(u_1 \le U_1^{t/\sum_{j=1}^n \log U_j} \le u_1 + \Delta_1, \dots, u_n \le U_n^{t/\sum_{j=1}^n \log U_j} \le u_n + \Delta_n)$$

for $\Delta_1, \ldots, \Delta_n > 0$. Thus the lemma holds for $\tau = 1$ by letting the $\Delta_i \to 0$.

For general $\tau > 0$ one applies the result above to $U_i = X_i/\tau$ and notes that $\sum \log X_i = s$ is equivalent to $\sum \log U_i = s - n \log \tau$. This ends the proof.

As a special result, we get the formula obtained by Lee [13] for simulating the conditional distribution of $\sum_{i=1}^{n} X_i$ given $\sum_{i=1}^{n} \log X_i = s$.

4.2 The log-linear law case

In the Introduction we indicated that the log-linear law NHPP does not admit simple pivotal statistics, thus making exact statistical analyses less straightforward. This fact is related to the fact that, in contrast to the result given in Lemma 1, there is apparently no simple direct way of sampling from the conditional distribution of uniforms X_1, \ldots, X_n on $[0, \tau]$ given $\sum_{j=1}^n X_j =$ s. The main reason for this is that there is no simple expression for the probability density of $\sum_{j=1}^{n} X_j$.

Inspired by, for example, Lockhart et al. [18] and Diaconis and Sturmfels [19], we shall use a slightly modified Gibbs sampler algorithm to simulate the desired samples.

The algorithm is based on the lemma below, which gives the conditional distribution of a pair (X_i, X_j) conditional on $\sum_{i=1}^n X_i = s$ and on the values of all X_k for $k \neq i, j$. Note first that the conditional distribution of X_1, \ldots, X_n given $\sum_{i=1}^n X_i = s$ is singular, so that in order to have a proper joint conditional density we have to leave out one variable, say X_n , and consider the conditional distribution of X_1, \ldots, X_{n-1} given $\sum_{i=1}^n X_i = s$.

Lemma 2 Let X_1, \ldots, X_n be independent and uniformly distributed on $[0, \tau]$. Then the conditional distribution of X_1 given

$$X_2 = x_2, X_3 = x_3, \dots, X_{n-1} = x_{n-1}, \sum_{i=1}^n X_i = s$$
 (10)

is given as

$$X_1 \sim \begin{cases} U[0,a] & \text{if } a \leq \tau, \\ U[a-\tau,\tau] & \text{if } a > \tau, \end{cases}$$
(11)

where $a = s - x_2 - x_3 - \cdots - x_{n-1}$.

Hence the conditional joint distribution of (X_1, X_n) given (10) can be represented by letting X_1 have distribution (11) while $X_n = a - X_1$.

Proof: The conditional distribution of X_1 given (10), which is given by a proper density by the comment preceding the lemma, must be the same as the conditional distribution of X_1 given $X_1 + X_n = s - x_2 - \cdots - x_{n-1}, X_2 = x_2, \ldots, X_{n-1} = x_{n-1}$. But by independence of the X_i , this distribution must be the same as the conditional distribution of X_1 given $X_1 + X_n = s - x_2 - \cdots - x_{n-1}, X_2 = \cdots - x_{n-1} \equiv a$. This distribution is straightforward to derive and is given by (11). This proves the lemma since the last part is trivial.

The algorithm is given next. Let M be a large number, depending on the application.

Gibbs sampler algorithm for conditional sampling of X_1, \ldots, X_n given $\sum_{i=1}^n X_i = s$, for i.i.d. $X_1, \ldots, X_n \sim U[0, \tau]$

1. Start with $x_i^0 = \frac{s}{n}$ for i = 1, ..., n; set m = 0.

- 2. Draw integers $1 \le i < j \le n$ randomly. Let $a = x_i^m + x_j^m$.
- 3. Draw

$$x_i^{m+1} \sim \begin{cases} U[0,a] & \text{if } a \leq \tau, \\ U[a-\tau,\tau] & \text{if } a > \tau. \end{cases}$$

Let

$$x_j^{m+1} = a - x_i^{m+1}.$$

4. If m < M, replace m by m + 1 and return to step 2.

The theory of the Gibbs sampler (see Casella and George [20], Gelfand and Smith [21]) guarantees that the distribution of the successive samples (x_1^m, \ldots, x_n^m) converges to the target distribution, whatever be the starting vector. In fact, the successive samples form a Markov chain, and the target distribution is the stationary distribution of this Markov chain. Thus, a certain number of "burn in" samples are needed before the samples can be taken to be from the correct distribution. It should also be noted that the successive samples are not independent, as they are in the power law case using Lemma 1.

5 Examples of goodness-of-fit tests for NHPP models

So far we have been concerned with the problem of conditional sampling of realizations for particular NHPP models, assuming that the test statistic Z is given. In practical goodness-of-fit testing, the first task might instead be to select a relevant test statistic Z = Z(T) with the ability to detect departures from an assumed parametric NHPP model. In this section we present some examples of such test statistics, all based on a useful transformation of the observed data to be presented first.

Consider an NHPP with intensity function $\lambda(t)$ and cumulative intensity function $\Lambda(t) = \int_0^t \lambda(u) du$. If T_1, T_2, \ldots are the successive event times of the NHPP, then $\Lambda(T_1), \Lambda(T_2), \ldots$ is a homogenous Poisson process with unit intensity. Suppose that the NHPP is observed on the time interval $[0, \tau]$. Then, conditional on $N(\tau) = n$, the transformed times $V_j = \Lambda(T_j)/\Lambda(\tau)$ for $j = 1, \ldots, n$ are distributed as the order statistic $U_{(1)}, \ldots, U_{(n)}$ of n i.i.d. variables with distribution U[0, 1].

If $\hat{\Lambda}(\cdot)$ is an estimate of $\Lambda(\cdot)$ based on the observation T in (1), then we shall define estimated transformed times $\hat{V}_1, \ldots, \hat{V}_n$ by $\hat{V}_j = \hat{\Lambda}(T_j) / \hat{\Lambda}(\tau)$. Since these appear as estimates of ordered uniform variables, it is tempting to use well-known goodness-of-fit tests for the uniform distribution to suggest goodness-of-fit test statistics for parametric NHPP-models based on the \hat{V}_j . This is, for example, the approach of Baker [2], Park and Kim [5], and Zhao and Wang [12].

5.1 The power law case

It was pointed out by Baker [2] that, for the power law process, the estimated transformed times \hat{V}_j , based on maximum likelihood estimates are pivots, in the sense that they have distributions which do not depend on the unknown parameters a, b.

To see this, note that in this situation we have

$$V_j = (T_j/\tau)^b; \quad j = 1, ..., n.$$
 (12)

The maximum likelihood estimate \hat{b} based on the observation T can be found by maximizing the likelihood (4), and is given by

$$\hat{b} = \frac{-n}{\sum_{j=1}^{n} \log T_j - n \log \tau} \tag{13}$$

(see Crowder et al. [14, p. 171]).

Now use the relation (12) and the representation $V_j = U_{(j)}$ where the $U_{(j)}$ are the orderings of i.i.d. variables in U[0, 1], to arrive at the representation $T_j = \tau U_{(j)}^{1/b}$. From this we can write the estimated transformed times \hat{V}_j as

$$\hat{V}_{j} = \left(\frac{T_{j}}{\tau}\right)^{\hat{b}} = U_{(j)}^{\frac{\hat{b}}{\hat{b}}} = U_{(j)}^{-n/\sum_{j=1}^{n} \log U_{j}}; \quad j = 1, ..., n,$$
(14)

where we have used (13) to get $\hat{b} = -nb/\sum_{j=1}^{n} \log U_j$. Now (14) does not involve the parameters a, b, and can therefore be simulated by means of a sample U_1, \ldots, U_n from U[0, 1]. The \hat{V}_j are hence pivots as announced.

Baker [2] studied a class of score tests based on the \hat{V}_j . Note that since the \hat{V}_j are pivots (in the power law case), we can compute the unconditional probabilities $P_{H_0}(Z \leq z_{obs})$ for test statistics based on the \hat{V}_j by simulation. But note that in general a goodness-of-fit statistic need not be a function of the \hat{V}_j , in which case we may need to base ourselves on conditional samples using Lemma 1, also for the power law case.

5.2 The log-linear law case

In this case the V_j are given by $V_j = (e^{bT_j} - 1)/(e^{b\tau} - 1)$ for j = 1, ..., n. The maximum likelihood estimate \hat{b} is found by solving the equation

$$\sum_{j=1}^{n} T_j + \frac{n}{b} - n\tau \left(\frac{1}{1 - e^{-b\tau}}\right) = 0 \tag{15}$$

with respect to b. This follows from (6) or from Crowder et al. [14, p. 167]. No closed form expression for \hat{b} exists, however.

The estimated transformed times V_j now become

$$\hat{V}_j = \frac{e^{\hat{b}T_j} - 1}{e^{\hat{b}\tau} - 1}; \ \ j = 1, ..., n.$$

In contrast to the power law case, the distributions of the \hat{V}_j depend on the parameter b and we can hence not simulate the distribution of a test statistic based on the \hat{V}_j unconditionally as for the power law case. Goodness-of-fit tests using the \hat{V}_j will therefore always be conditional tests in our approach.

5.3 Test statistics

We present as illustration five different test statistics based on the \hat{V}_j . For simplicity we put $N(\tau) = n$ in the formulae.

Laplace statistic (Zhao and Wang [12])

$$L = \sqrt{\frac{12}{n}} \sum_{j=1}^{n} (\hat{V}_j - \frac{1}{2}).$$
(16)

This is a two-sided test statistic and the null hypothesis of NHPP is rejected for either too small or too large values.

Greenwood statistic (Baker [2])

$$G = \sum_{j=1}^{n+1} (\hat{V}_j - \hat{V}_{j-1})^2, \qquad (17)$$

where $\hat{V}_{n+1}=1$ and $\hat{V}_0=0$. This is a two-sided test statistic and the null hypothesis of NHPP is rejected for either too small or too large values.

Modified Cramer-von Mises statistic (Park and Kim [5])

$$W^{2} = \sum_{j=1}^{n} [\hat{V}_{j} - \frac{(2j-1)}{2n}]^{2} + \frac{1}{12n}.$$

Modified Anderson-Darling statistic (Park and Kim [5])

$$A^{2} = -\{\sum_{j=1}^{n} (2j-1) [\log \hat{V}_{j} + \log(1-\hat{V}_{n+1-j})]\}/n - n.$$

Modified Kolmogorov-Smirnov statistic (Park and Kim [5])

$$D = \max \ [D^+, D^-],$$

where

$$D^+ \equiv \max_{1 \le j \le n} (\frac{j}{n} - \hat{V}_j), \ D^- \equiv \max_{1 \le j \le n} (\hat{V}_j - \frac{(j-1)}{n})$$

The last three test statistics are versions of traditional goodness-of-fit statistics for distributions. The corresponding tests reject the null model for large values of the statistic.

6 Processes with failure censoring

Assume in this section that a repairable system is observed until a prespecified number of events, say n. The following modifications need to be done in the methods considered so far, which all assume time censoring.

First, the data can now be described simply as $T = (T_1, \ldots, T_n)$, while the likelihood function (2) becomes

$$\sum_{j=1}^{n} \log \lambda(T_j) - \int_0^{T_n} \lambda(u) du.$$
(18)

For the power law case this becomes

$$l_p(a,b) = n(\log a + \log b) + (b-1)\sum_{j=1}^n \log T_j - aT_n^b$$

which should be compared to (4). It is hence seen that the sufficient statistic can be given as

$$S_p(T) = (T_n, \sum_{j=1}^{n-1} \log T_j).$$

Similarly, it can be shown that the sufficient statistic in the log-linear law case is

$$S_{\ell}(T) = (T_n, \sum_{j=1}^{n-1} T_j).$$

Thus, all the information on the parameters are now in the last failure time, T_n , and in the sum of the (logs) of the (n-1 first) observations.

The conditional simulation recipes from Section 4 can be modified using the result (e.g. Ross [17]) that, conditionally on $T_n = t_n$, the event times $T_1, T_2, \ldots, T_{n-1}$ of an HPP(1) are distributed as the ordering of n-1 i.i.d. random variables distributed as $U[0, t_n]$.

Comparing with the corresponding result used in Section 4, this means that conditional simulations given the sufficient statistics can be performed as for the time-censored case, with τ replaced by t_n , and with the number of simulated values reduced from n to n-1.

As regards maximum likelihood estimation of b, looking at the likelihood function (18), it is seen that we can use the same expression for the maximum likelihood estimator \hat{b} as for the time-censored cases, just replacing τ by t_n .

Consider finally the definition and computation of the transformed times V_j and \hat{V}_j . These are now defined as

$$V_j = \Lambda(T_j) / \Lambda(T_n); \quad j = 1, \dots, n-1,$$
(19)

which for the power law case lead to $V_j = (T_j/T_n)^b$. It can be shown that the estimated ones, given by $\hat{V}_j = (T_j/T_n)^{\hat{b}}$, can be represented by an expression similar to (14), namely

$$\hat{V}_j = \left(\frac{T_j}{T_n}\right)^{\hat{b}} = U_{(j)}^{\hat{b}/b} = U_{(j)}^{-n/\sum_{j=1}^{n-1} \log U_j}; \quad j = 1, ..., n$$

for a sample U_1, \ldots, U_{n-1} from U[0, 1].

The \hat{V}_j in the log-linear law case are found from (19) to be

$$\hat{V}_j = \frac{e^{bT_j} - 1}{e^{\hat{b}T_n} - 1}; \ j = 1, ..., n - 1,$$

where \hat{b} is the maximum likelihood estimator obtained by solving (15) for b when τ is replaced by T_n .

7 Application and simulations

7.1 Data example (Leitch [1])

We apply the results to the data from Leitch [1], presented in the Introduction. Since there is no information beyond the last failure time 8245, we assume failure censoring in the following.

Let us first check whether the data are consistent with a power law NHPP, using some of the test statistics presented in Section 5. Recall that for the power law, we are able to simulate unconditionally the distribution of all statistics based on the \hat{V} .

We first apply the Greenwood statistic (17) and calculate its value $g_{obs} = 0.1263$ for the observed data. Then we simulate the distribution of G. This is done by simulating 100.000 samples of $\{\hat{V}_j\}$ from (14) and each time computing the statistic G. We then get the upper histogram in Figure 1, where we have also marked the observed value g_{obs} based on the original data. The resulting *p*-value is found to be 0.0487 (corresponding to twice the tail to the left of the observed value since the test based on G is two-sided). Hence the Greenwood statistic implies evidence against the power law assumption for the reliability growth data.

We now do the corresponding analyses using, for illustration, the Cramer von Mises test and the Kolmogorov-Smirnov test (see Section 5). The results are given, respectively, in the middle and lower plots in Figure 1, together with the calculated test statistics and corresponding p-values.

Neither of the two last tests give any evidence against power law NHPP, and the observed p-values in fact deviate remarkably from the p-value of the Greenwood test. This leads to the question of why this difference from the Greenwood test is observed? In order to answer this question we shall take a closer look at the test statistics, see the discussion at the end of this subsection.

Let us next check whether the data are consistent with a log-linear law NHPP. We did 1.000.000 iterations using the Gibbs algorithm of Section 4, each time computing the value of the test statistic. To have comparable scales on the plots of power law and log-linear law, we used every 10th value of the simulated test statistics to draw the histogram. The resulting histograms are in Figure 2. Note, however, that the simulations are now conditional on the sufficient statistic.

The most remarkable difference to the power law case is that the (conditional) *p*-value for the Greenwood statistic is higher for the log-linear law model, thus not suggesting departure from this NHPP model. The (conditional) *p*-values for the Cramer von Mises and Kolmogorov-Smirnov statistics are lower than for the power law case, but still high.

Figure 3 shows the transformed times \hat{V}_j for the power law and the loglinear law, respectively. Comparing these with samples from i.i.d. uniforms, the points are remarkably regularly positioned. The Greenwood statistic can be shown to have its minimum value when the \hat{V}_j are evenly spread out. This probably explains the low observed value of G for these data. In fact,



Figure 1: Reliability growth data: Simulated distributions of test statistics for power law NHPP (unconditional distributions). Upper curve is for Greenwood statistic G; middle curve is for Cramer von Mises statistic W^2 ; lower curve is for Kolmogorov-Smirnov statistic D. Observed values of test statistics are marked with vertical bar.



Figure 2: Reliability growth data: Simulated distributions of test statistics for log-linear NHPP (conditional distributions). Upper curve is for Greenwood statistic G; middle curve is for Cramer von Mises statistic W^2 ; lower curve is for Kolmogorov-Smirnov statistic D.Observed values of test statistics are marked with vertical bar.



Figure 3: Reliability growth data: Transformed times \hat{V} computed from the power law (upper plot) and the log-linear model (lower plot).

Figure 3 also indicates why the value of G is higher in the log-linear case, since the regularity is apparently a bit less pronounced there. On the other hand, the Cramer von Mises and Kolmogorov-Smirnoff statistics, by their definition, will be small when the \hat{V}_j are regularly positioned, since then the empirical distribution of the \hat{V}_j is close to the distribution of U[0, 1]. This is believed to explain the high *p*-values for the corresponding tests.

7.2 A simulation study

The somewhat unexpected behavior of the Cramer von Mises test in situations as the one considered in the above example was noted by Baker [2], who concluded that this test has low power against certain violations of the NHPP property itself (here revealed as the "regularity" mentioned in the example). This behavior motivated the small simulation study presented in the following.

Table 1 shows the result of a simulation of several goodness-of-fit tests when the alternative is a non-NHPP process. 100.000 data sets are simulated, with failures following a renewal process until n = 10 failures. The distribution of the inter-failure times is Weibull with scale parameter 1 and varying shape parameter α , i.e. a distribution with survival function $\exp\{-t^{\alpha}\}$. Table 1 shows the relative number of rejections (power) for testing the null hypothesis that the data come from a power law NHPP using the tests from Section 5. We first note that all tests have power close to the chosen significance level 0.05 when $\alpha = 1$, i.e. when we have an HPP and hence a power law NHPP. This should of course be so since we consider exact tests. It is remarkable, however, that all the tests, except the Greenwood test, have power decreasing with α , which leads to powers less than 0.05 for alternatives with $\alpha > 1$. (This problem is already noted in [2] for the Cramer von Mises

	Laplace	Greenw	Cramer v M	And-Darl	Kolm-Smir
α	L	G	W^2	A^2	D
0.25	0.4787	0.8427	0.6841	0.8379	0.7063
0.50	0.2726	0.4543	0.3430	0.4625	0.3713
0.75	0.1250	0.1396	0.1321	0.1678	0.1444
1	0.0500	0.0496	0.0497	0.0506	0.0502
1.25	0.0189	0.0804	0.0208	0.0151	0.0178
1.50	0.0070	0.1708	0.0081	0.0046	0.0055
1.75	0.0027	0.3074	0.0034	0.0014	0.0018
2	0.0011	0.4636	0.0013	0.0006	0.0005

Table 1: Power of goodness-of-fit tests of the null hypothesis of power law NHPP for n = 10 failure censored times and significance level 0.05, when data come from a Weibull renewal process with shape parameter α .

test). The Greenwood test is on the other hand able to detect alternatives with $\alpha > 1$, and is together with the Anderson-Darling test the best to detect alternatives with $\alpha < 1$.

Table 2 is included to show a different behavior of the test statistics when the alternative to power law NHPP is an NHPP with another parametric form, here with intensity function $\lambda(t) = \alpha\beta/(1+\beta t)$ for parameters $\alpha, \beta > 0$. This is called the Musa-Okumoto model (see e.g. [12]). Again 100.000 data sets are simulated. We first note that all tests have a power which decreases with α . This is intuitively so because the process comes closer and closer to an HPP as α increases. Indeed, we can see that the powers are close to 0.05 when $\alpha = 1000$, which is presumably a case close to an HPP. A remarkable difference from Table 1 is of course that the Greenwood statistic now has very low power throughout, while the three last test statistics apparently are the best ones to detect alternatives to power law NHPP which are still NHPPs, but have a different intensity function.

7.3 Exact tests versus asymptotic tests

As already mentioned, a main motivation for deriving and using exact tests comes from the fact that critical values for tests based on asymptotics may be misleading. We may exemplify this by considering the goodness-of-fit test based on the Laplace statistic considered by Zhao and Wang [12]. The asymptotic test rejects the null hypothesis of a power law NHPP if $|L| > z_{\alpha/2}/2$ (see [12]). This is valid for large *n* only, however. We found in fact by simulation that for data from an HPP (and hence a power law NHPP)

	Laplace	Greenw	Cramer v M	And-Darl	Kolm-Smir
α	L	G	W^2	A^2	D
0.1	0.2908	0.0625	0.3884	0.3314	0.3156
0.5	0.2722	0.0615	0.3648	0.3080	0.2979
1	0.2387	0.0595	0.3211	0.2654	0.2674
2	0.1656	0.0555	0.2324	0.1847	0.1998
3	0.1242	0.0519	0.1780	0.1399	0.1574
5	0.0812	0.0503	0.1196	0.0937	0.1108
10	0.0579	0.0494	0.0817	0.0668	0.0780
1000	0.0511	0.0490	0.0518	0.0512	0.0524

Table 2: Power of goodness-of-fit tests of the null hypothesis of power law NHPP for n = 10 failure censored times and significance level 0.05, when data come from the Musa-Okumoto NHPP with parameters $\beta = 1$ and varying parameter α .

with n = 10 we obtain a power of 0.14 when the nominal level was set to $\alpha = 0.05$.

We also computed the Laplace-statistic for the reliability growth data considered in the example of Section 7.1, and simulated the distribution of L for a power law NHPP with n = 10, in order to compute the *p*-value (see Figure 4). The figure shows that the distribution of L is far from normal for n = 10 (skew to the right). This explains the large difference between the nominal level and actual rejection probability when data come from an HPP, as reported above. In fact, for n = 10 and significance level 0.05 the asymptotic Laplace test would reject for |L| > 0.98, which looks unreasonable from Figure 4. Instead, cutting off 2.5% on each side of the histogram in Figure 4 gives the correct critical values -1.27 and 0.68.

8 Concluding remarks

The present paper advocates the use of exact goodness-of-fit tests for parametric NHPP models and suggests that such tests can be derived by conditioning test statistics on a sufficient statistic under the null hypothesis model. Concrete recipes for how to do this are given for the power law and log-linear law cases. For other models, it may or may not be straightforward to follow similar procedures. We give below some relevant references to the literature.

Lee [13] considered a parametric NHPP model with both the power law



Figure 4: Reliability growth data: Simulated distribution of the Laplace test statistic L for power law NHPP, with observed value of L marked with vertical bar.

and log-linear law as special cases, given by the intensity function

$$\lambda_L(t) = abt^{b-1}e^{ct}.$$

This intensity function can be both increasing, decreasing or a combination of the two. Lee's idea was that, by adding a parameter to the power law or log-linear law models, a more flexible model could be obtained, with an important feature being to use the new model to derive goodness-of-fit tests for the power law and log-linear law cases, by testing respectively the null hypotheses c = 0 and b = 1.

However, one may also want to test the null hypotheses that λ_L is the true intensity function. In this case, the sufficient statistic for data T as in (1) is

$$S_L(T) = (N(\tau), \sum_{j=1}^{N(\tau)} T_j, \sum_{j=1}^{N(\tau)} \log T_j).$$

Following the same kind of reasoning as in Section 4, the challenge is now to sample n uniforms on $[0, \tau]$ given both their sum and the sum of the logs. We will not go into details here about how this can be done, but the case is quite similar to the case of conditional sampling in gamma models, as studied by Lockhart et al. [18]. The difference is that, while we consider conditional samples of uniforms, they consider conditional samples from exponential distributions.

The Gibbs sampling method used in Section 4 will presumably work in rather general situations. It should be noted that parametric NHPP models

usually contain a scale parameter, like the parameter a in the power law case. The idea of first conditioning on $N(\tau) = n$ should then be recommended generally, since it leads to the elimination of this scale parameter.

More general approaches to conditional sampling given sufficient statistics are found in Lindqvist and Taraldsen [22] and Lockhardt et al. [18]. The methods given there may well be applied to a large class of NHPP models. Note also, that Lindqvist and Taraldsen [22] study the problem of conditional sampling of uniforms given their sum, using another type of algorithm than the one considered in the present paper.

We would finally like to mention that, if one relaxes the requirement of exactly computed p-values, but finds that asymptotic theory does not give satisfactory answers, then parametric bootstrapping is a powerful tool. The basic idea is then to simulate samples from the null hypothesis model by substituting parameter values estimated under the null hypothesis assumption. A short discussion of this is given in Lockhart et al. [18].

References

- Leitch RD. Reliability analysis for engineers. New York: Oxford University Press, 1995.
- [2] Baker RD. Some new tests of the power law process. *Technometrics* 1996;38:256-265.
- Bain LJ, Engelhardt M. Statistical analysis of reliability and life-testing models (2nd ed.) New York: Marcel Dekker, 1991.
- [4] Crow LH. Reliability analysis for complex, repairable systems. In *Re-liability and biometry*. Proschan F, Serfling RJ, editors. Philadelphia, Pennsylvania: SIAM;1974. p. 379-410.
- [5] Park WJ, Kim YG. Goodness-of-fit tests for the power-law process. *IEEE Transactions On Reliability* 1992;41:107-111.
- [6] Lindqvist BH, Elvebakk G, Heggland K. The trend-renewal process for statistical analysis of repairable systems. *Technometrics* 2003;45:31-44.
- [7] Kumar U, Klefsjö B. Reliability analysis of hydraulic systems of LHD machines using the power law process model. *Reliability Engineering and* System Safety 1992;35:217-224.
- [8] Gaudoin O. CPIT goodness-of-fit tests for the power-law process. Communications in Statistics, Theory and Methods 1998;27(1):165-180.

- [9] Gaudoin O. CPIT Goodness-of-fit tests for reliability growth models. In: Statistical and Probabilistic Models in Reliability, D.C. Ionescu and N. Limnios eds, Birkhäuser, Boston, 1999, 27-37.
- [10] Crétois E, El Aroui MA, Gaudoin O. U-plot method for testing the goodness-of-fit of the power-law process. *Communications in Statistics*, *Theory and Methods* 1999;28(7):1731-1747.
- [11] Gaudoin O, Xie M, Yang N. A simple goodness-of-fit test for the powerlaw process based on the Duane plot. *IEEE Transactions on Reliability* 2003;52(1):69-74.
- [12] Zhao J. and Wang J. A new goodness-of-fit test based on the Laplace statistic for a large class of NHPP models. *Communications in Statistics, Simulation and Computation* 2005;34:725-736.
- [13] Lee L. Testing adequacy of the Weibull and log linear rate models for a Poisson process. *Technometrics* 1980;22(2):195-199.
- [14] Crowder MJ, Kimber AC, Smith RL, Sweeting TJ. Statistical analysis of reliability data. London: Chapman & Hall, 1991.
- [15] Meeker WQ, Escobar LA. Statistical methods for reliability data. New York: Wiley, 1998.
- [16] Casella G, Berger RL. Statistical inference (2nd ed.). Pacific Grove, Calif.: Duxbury, 2002.
- [17] Ross SM. Introduction to probability models (9th ed.). Amsterdam: Elsevier, 2007.
- [18] Lockhart RA, O'Reilly FJ, Stephens MA. Use of the Gibbs sampler to obtain conditional tests, with applications. *Biometrika* 2007;94:992-998.
- [19] Diaconis P, Sturmfels B. Algebraic algorithms for sampling from conditional distributions. Annals of Statistics 1998;26(1):363-397.
- [20] Casella G, George EI. Explaining the Gibbs sampler. *The American Statistician* 1992;46:167-174.
- [21] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990;85:398-409.

[22] Lindqvist BH, Taraldsen G. Monte Carlo conditioning on a sufficient statistic. *Biometrika* 2005;92:451-464.

Bo Henry Lindqvist Department of Mathematical Sciences Norwegian University of Science and Technology N-7491 Trondheim, Norway Email: bo@math.ntnu.no